

Penerapan Regresi Logistik Biner Terboboti Geografi dengan Pembobot *Fixed Bi-Square*

(Geographically Weighted Binary Logistic Regression with Fixed Bi-Square Weight)

Tuti Purwaningsih

Universitas Islam Indonesia

Jl Kaliurang KM 14.5, Yogyakarta

tuti.purwaningsih@uii.ac.id

ABSTRACT

Geographically weighted regression (GWR) is one of model in spasial regression. It used when the spatial regression had heteroschedasticity. Each location will have different model based on the significant coefficient in every explanatory variable at each location. Then if the respon variable is binary data so the model will be mentioned as Geographically Weighted Binary Logistic Regression. The study has aim to applicate it in real data. The Data taken from PODES which consist of one respon variable with binary type and three explanatory variable with ratio type. The respon variable is poverty category for each municipality and the independet variable is percentage of trade village, percentage of service village and ratio of school pervillage. This research gives Global Model and Local Model. At Global Model only two explanatory variables which has significant influence to the poverty category with AIC value about 122,563. At Local Model gives different result based on the significant coefficient in each municipality.

Keywords : GWR, Geographically Weighted Binary Logistic Regression, *Fixed Bi-Square*, Global Model, Local Model

ABSTRAK

Regresi terboboti geografi adalah salah satu model yang digunakan dalam regresi spasial. Model tersebut digunakan ketika regresi spasial mengalami masalah heteroskedastisitas. Setiap lokasi akan memiliki model yang berbeda-beda tergantung dari signifikansi koefisien regresi yang ada pada setiap variable disetiap lokasi. Kemudian jika tipe data dari variabel respon adalah biner maka akan disebut sebagai regresi logistik biner terboboti geografi. Penelitian ini bertujuan untuk menerapkan model tersebut pada kasus riel. Data yang digunakan diambil dari survei PODES yang terdiri dari satu variabel respon bertipe biner dan tiga variabel penjelas bertipe rasio. Variabel respon tersebut adalah kategori kemiskinan pada setiap kabupaten dan variabel penjelasnya adalah persentase desa perdagangan, persentase desa jasa dan rasio sekolah perdesa. Dari penelitian ini menghasilkan dua jenis model yaitu model global dan model lokal. Pada model global hanya dua variabel penjelas yang signifikan berpengaruh terhadap kategori kemiskinan di Pulau Jawa dengan nilai AIC sebesar 122.563. Pada Model Lokal menghasilkan hasil yang berbeda tergantung dari koefisien yang signifikan pada setiap kabupaten.

Keywords : GWR, Regresi Logistik Biner Terboboti Geografi, *Fixed Bi-Square*, Model Global, Model Lokal

Pendahuluan

Analisis regresi mengalami perkembangan yang cukup luas sekarang ini, salah satu bidang yang cukup berkembang adalah spasial. Analisis spasial, memperhatikan lokasi, keterkaitan antar lokasi dan jarak antar lokasi yang satu dengan lokasi lainnya. Pada umumnya, analisis spasial mengasumsikan bahwa antar objek (antar lokasi) tidak bebas stokastik, karena mereka saling berkaitan.

Hal ini dikenal dengan adanya korelasi spasial yang biasa diukur dengan menggunakan Indeks Moran. Dalam mendeteksi parameter parameter yang berkaitan dengan lokasi, juga bisa mengidentifikasi tingkat asosiasi suatu parameter antar lokasi, misalkan asosiasi beberapa spesies dengan habitatnya. Setelah mengetahui besaran korelasi spasial maupun asosiasi suatu parameter, analisis selanjutnya adalah mengukur besaran pengaruh variabel independen terhadap variabel dependen yang dikenal dengan regresi spasial.

Penelitian regresi spasial terus berkembang dan saat ini dikenal juga berbagai macam model regresi spasial antara lain Regresi Logistik Spasial, Regresi Lag Spasial (lag y atau lag x), *Spatial Error Model (SEM)*, *Spatial Autoregression Model (SAR)*, serta *Geographically Weighted Regression (GWR)*, dll.

Pada penelitian ini difokuskan untuk menerapkan model Regresi Terboboti geografi atau yang biasa disebut *Geographically Weighted Regression (GWR)* dengan menggunakan data biner

sebagai variabel responnya yang disebut sebagai Regresi Logistik Biner Terboboti Geografi (RLBTG) dengan studi kasus yang digunakan mengenai data kemiskinan di pulau Jawa.

Tujuan Penelitian

1. Menerapkan regresi logistik ordinal terboboti geografi
2. Mengetahui pembobot terbaik dalam model regresi logistik ordinal terboboti geografi pada kasus kemiskinan
3. Menganalisis model global dan model lokal yang dihasilkan

Metode Penelitian

Penerapan analisis regresi spasial yang akan kita pelajari adalah analisis regresi terboboti geografi (RTG) dengan studi kasus yang diambil adalah kemiskinan kabupaten di pulau Jawa.

1. Data

Data yang digunakan berupa data PODES yang terdiri dari faktor-faktor yang diduga mempengaruhi kemiskinan di Pulau Jawa pada setiap kabupatennya (108 kabupaten). Terdiri dari satu variabel respon dan tiga variabel penjelas.

Variabel respon berupa data kategori yaitu miskin dan tidak miskin, jika kategori miskin maka nilainya 1 dan jika kategori tidak miskin maka nilainya 0. Tiga variabel penjelas yang digunakan, sebagai berikut:

Tabel 1. Nama Variabel dalam Model

Variabel	Definisi
Y	Kategori kemiskinan (1=miskin, 0=tidak miskin)
X1	Persentase Desa Perdagangan besar/eceran
X2	Persentase Desa jasa
X3	Rasio sekolah/desa

Akan dianalisis apakah tiga variabel penjelas tersebut merupakan variabel yang mempengaruhi kemiskinan kabupaten-kabupaten di Pulau Jawa ataukah tidak. Analisis yang digunakan adalah metode Regresi Terboboti Geografis, sehingga setiap kabupaten/kota di Pulau Jawa mempunyai model regresi terboboti tersendiri untuk menggambarkan kondisi kemiskinannya masing-masing.

2. Metodologi

Langkah-langkah yang dilakukan untuk analisis Regresi Terboboti Geografi (RTG) untuk logistik biner, adalah sebagai berikut:

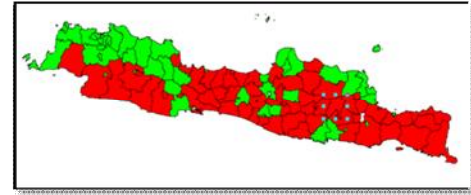
1. Memastikan bahwa data dari variable respon adalah data biner
2. Dalam langkah kedua sebenarnya kita perlu menguji apakah data dari variable respon mengalami masalah heterokedastisitas (Uji Breusch Pagan) atau tidak, akan tetapi karena data variable respon berupa data biner, maka kita tidak perlu mengujinya karena nilai errornya berkisar antara 0 dan 1 (berupa data kualitatif)
3. Melakukan analisis regresi terboboti geografi dengan tiga variable respon

4. Membandingkan keempat metode pembobotan untuk mendapatkan model terbaik

5. Menginterpretasikan hasil output

Pembahasan

Berikut adalah output yang dihasilkan berdasarkan kategori kemiskinan setiap kabupaten di Pulau Jawa.



Keterangan:

	0 Cenderung tidak miskin
	1 Cenderung Miskin

Gambar 1. Peta Tematik Kategori Kemiskinan Kabupaten

Berdasarkan Gambar 1 dapat dikatakan bahwa mayoritas kabupaten masuk dalam kategori cenderung miskin yang tersebar di 4 provinsi di Pulau Jawa yaitu Jawa Barat, Jawa tengah, Yogyakarta dan Jawa Timur. Secara visual dapat dikatakan bahwa kemiskinan antar kabupaten saling terkait antara lokasi yang bertetangga, hal ini mengindikasikan adanya korelasi spasial kemiskinan di pulau Jawa.

1. Model Regresi Global

Setelah melakukan analisa regresi logistik biner dengan GWR, didapatkan Model regresi global sebagai berikut:

Tabel 2. Estimasi Model RLBTG

Variabel	Koef	SE	Z (Koef/SE)	Exp (Koef)
Intersep	1.529551	0.545	2.806	4.616
X1	2.074296	1.979	1.047	7.958
X2	-4.473083	2.448	-1.826	0.011
X3	-0.2015131	0.125	-1.608	0.817

Hasil yang terdapat pada Tabel 2 merupakan model regresi logistik biner yang berlaku untuk semua lokasi (bersifat Global), dari hasil estimasi pada table 1 maka dapat dipaparkan model RLBTG sebagai berikut :

$$\text{Log} \frac{p}{1-p} = b_0 + b_1 X_1 - b_2 X_2 - b_3 X_3 + e$$

$$\text{Log} \frac{p}{1-p} = 1,529551 + 2,07496 X_1 - 4,473083 X_2 - 0,2015131 X_3 + e$$

2. Pengujian dan interpretasi parameter model regresi global

$$b_1 = 2,074296$$

Memiliki nilai Statistik wald $1.047668 < t$ tabel = 1.45 (alpha =15%, dbg=104)

Sehingga dapat disimpulkan bahwa variabel X1 tidak signifikan pengaruhnya terhadap kemiskinan di Jawa secara Global dengan tingkat kepercayaan 85%.

$$b_2 = -4,473083$$

Memiliki nilai Statistik wald $|-1.826556| > t$ tabel = 1.45 (alpha =15%, dbg=104)

Sehingga dapat disimpulkan bahwa variabel X2 signifikan pengaruhnya terhadap kemiskinan di Jawa secara Global dengan tingkat kepercayaan 85%.

$$b_3 = -0,2015131$$

Memiliki nilai Statistik wald $|-1.608815| > t$ tabel = 1.45 (alpha =15%, dbg=104)

Sehingga dapat disimpulkan bahwa variabel X3 signifikan pengaruhnya terhadap kemiskinan di Jawa secara Global dengan tingkat kepercayaan 85%

3. Perbandingan keempat metode Kernel untuk mendapatkan model terbaik

Untuk memilih metode kernel terbaik, dilakukan pembuatan model untuk masing-masing pembobot untuk mendapatkan nilai AIC dari pembobot. Pembobot dengan nilai AIC paling kecil digunakan sebagai pembobot terbaik untuk membangun model. Pada Tabel 2 dapat dilihat hasil AIC masing-masing pembobot.

Tabel 3. Perbandingan AIC

Pembobot kernel	AIC(global)
adaptive bi-square	126,959861
adaptive Gaussian	129,647688
fixed bi-square	122,563356
fixed Gaussian	124,342202

Pembobotan yang akan digunakan sebagai pembobot model adalah pembobotan yang menghasilkan nilai AIC terkecil yaitu *Fixed- Bisquare*. Nilai AIC terkecil yang dihasilkan yaitu dari metode *Fixed- Bisquare* (122,563356) dan selisih dengan metode lainnya lebih besar dari 2, jadi dipilih metode *Fixed- Bisquare*.

4. Model Lokal

Berdasarkan hasil analisa model lokal, di ambil salah satu contoh untuk wilayah Jember.

Tabel 4. Hasil Model Lokal Kab. Jember

Area_num	est_X1	t_X1	est_X2	t_X2	est_X3	t_X3
Jember	2,491	1,044	-4,318	-1,556	-0,190	-1,244

Berdasarkan hasil model lokal daerah Jember, didapatkan X1 dan X3 tidak berpengaruh signifikan terhadap kemiskinan

di Jember. Variabel X2 (% desa jasa) yang berpengaruh signifikan.

$$\text{Log} \frac{\hat{p}}{1-\hat{p}} = 1,707 - 4,3189x_2$$

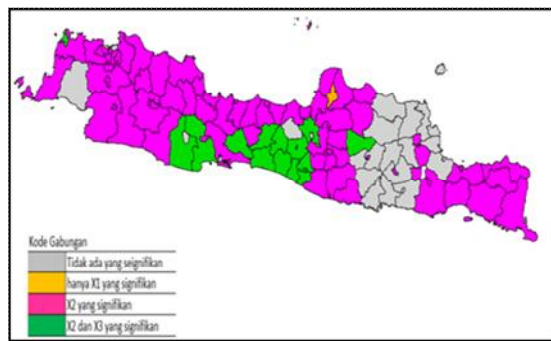
Interpretasi

Dilihat dari nilai odds ratio untuk koefisien x2 yaitu $e^{-4,3189} = 0,013314$

Maka peluang daerah Jember untuk miskin sebesar 0,013314 kali peluang daerah Jember untuk tidak miskin jika persentase desa jasa meningkat sebanyak 1% dari persentase data sebelumnya

Interpretasi untuk daerah lainnya tergantung dari model lokal masing-masing. Dilihat dari variabel mana yang signifikan, kemudian untuk mengetahui besarnya pengaruh maupun kontribusi variabel tersebut, dihitung terlebih dahulu nilai odds rasionya.

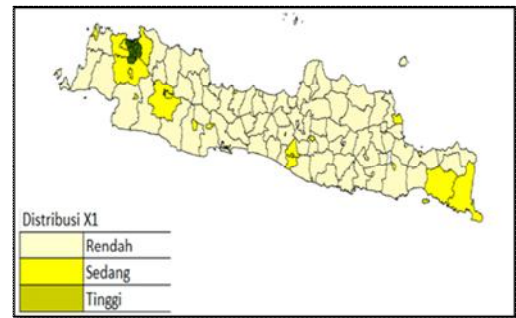
Analisa lebih lanjut mengenai variabel-variabel penjelas pada masing-masing lokasi, dapat dilihat secara keseluruhan berdasarkan beberapa peta distribusi di bawah ini:



Gambar 2. Distribusi Signifikansi X1, X2, X3

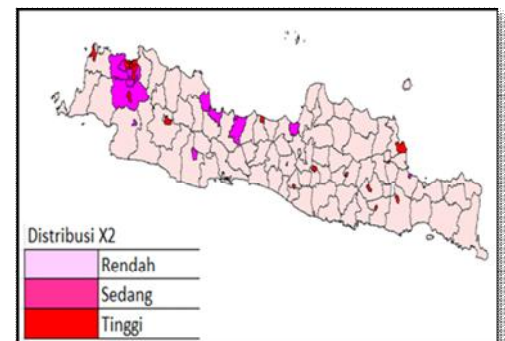
Berdasarkan peta tematik pada Gambar 2 terlihat bahwa X2 lebih

banyak yang signifikan pada banyak kabupaten di pulau Jawa, sedangkan X1 yang paling sedikit.



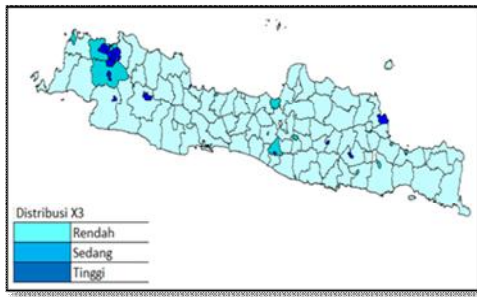
Gambar 3. Distribusi Signifikansi X1

Berdasarkan peta pada Gambar 3 terlihat bahwa wilayah Jakarta memiliki persentase desa perdagangan yang lebih tinggi di bandingkan wilayah lainnya dan secara umum persentase desa perdagangan mayoritas wilayah masih rendah.



Gambar 4. Distribusi Signifikansi X2

Berdasarkan peta pada Gambar 4, menunjukkan bahwa sebagian besar wilayah yang memiliki persentase sedang dan tinggi, mayoritas ada di daerah Jakarta dan Jawa Barat. Dan sebagian kecil ada di Jawa Tengah kemudian untuk Jawa Timur terjadi sedikit ketimpangan dimana persentase desa jasa hanya ada tinggi dan rendah, hal ini menunjukkan bahwa adanya kecenderungan korelasi spasial negatif antar daerah di Jawa Tengah dan Jawa Timur.



Gambar 5. Distribusi Signifikansi X3

Berdasarkan peta distribusi variable rasio sekolah/desa pada Gambar 5 menunjukkan bahwa daerah yang memiliki persentase paling tinggi berada pada wilayah Jakarta dan sekitarnya sedangkan daerah pada provinsi Jawa Tengah dan Jawa Timur rata-rata memiliki persentase yang rendah untuk variabel ini.

Kesimpulan

Dari analisis yang sudah dilakukan, maka dapat kita simpulkan bahwa:

1. Pada penelitian ini berhasil menerapkan model Regresi Logistik Biner Terboboti Geografi untuk kasus kemiskinan di Pulau Jawa
2. Dari tiga variable independen, yang berpengaruh signifikan pada model Global adalah variable X_2 dan X_3
3. Model Global memiliki 2 variabel independen yang signifikan dari tiga variabel yang diikutkan kedalam model, sedangkan pada model lokal menghasilkan signifikansi yang berbeda-beda dari setiap variabel independen pada setiap kabupaten.

Pustaka

- Agresti A. 2007. *Categorical Data Analysis*. New Jersey: John Wiley and Sons.
- Aidi MN, Purwaningsih T. 2013. Modelling Spatial Ordinal Logistic Regression and The Principal Component to Predict Poverty Status of Districts in Java Island”, *International Journal of Statistics and Application* 3(1):1-8
- Anselin, L. 1998. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.
- Brunsdon, C., Fotheringham A.S., Charlton M. 1999. *Some notes on parametric significance tests for geographically weighted regression*, *Journal of Regional Science*, Vol. 39, No 3, 497-524
- Cressie Noel A C. 1993. *Statistics for Spatial Data*. Revised Edition. John Wiley & Sons, Inc.
- Draper, N.R, & Smith, H. 1998. *Applied Regression Analysis 3rd Edition*. Canada : John Wiley and Sons, Inc.
- Fotheringham A.S., Brunsdon C., Chartlon M. (2002). *Geographically Weighted Regression, the analysis of spatially varying relationships*, John Wiley and Sons, LTD.
- Fotheringham AS, Rogerson PA. 2009. *Spatial Analysis*. London: Sage Publications, Inc.
- Garson GD. *Logistic Regression*. [link] <http://www.chass.ncsu.edu> [11 April 2015].
- Haining Robert. 2004. *Spatial Data Analysis Theory and Practice*. Cambridge University Press.
- Hosmer DW, Lemeshow S. 2000. *Applied Logistic Regression Second Edition*. New York : John Wiley and Sons.
- Lee J, Wong DWS. 2001. *Statistic for Spatial Data*. New York : John Wiley & Sons, Inc.
- Lee J, Wong DWS. 2001. *Statistical Analysis ArchView GIS*. New York: John Wiley & Sons, Inc.
- Leung Y., Mei Chang L., and Zhang Wen X. 2000. *Statistical Test for*

Spatial Non Stationarity based on the geographically Weighted Regression Model, Environment and Planning, vol. 32, 9-32.

Purwaningsih T, et.al. 2015. Comparison of Unifrom and Kernel Gaussian Weight Matrix in Generalized Spatial Panel Data Model. Open Journal of Statistics 5, 90-95.

Ward MD, Gleditsch KS. 2008. *Spatial Regression Models*. Los Angeles: Sage Publications, Inc.